

# A circle of interacting servers; spontaneous collective behavior in case of large fluctuations

E.A. Pechersky and N.D. Vvedenskaya

Dobrushin Mathematical Laboratory,

Institute for Problems of Information Transmission, Russian Academy of Sciences,

B. Karetnyi per., GSP-4 Moscow 101447, Russia

pech@iitp.ru, ndv@iitp.ru

## Abstract

We consider large fluctuations, namely overload of servers, in a network with dynamic routing of messages. The servers form a circle. The number of input flows is equal to the number of servers, the messages of any flow are distributed between two neighboring servers, upon its arrival a message is directed to the least loaded of these servers. Under the condition that at least two servers are overloaded the number of overloaded servers in such network depends on the rate of input flows. In particular there exists critical level of input rate that in case of higher rate most probable that all servers are overloaded.

**Keywords:** Large Deviation Principle, Queuing Networks, Dynamic Routing.

## 1 Introduction

This work presents an effect in a network with interacting servers that can be called a spontaneous collective behavior in case of large fluctuations.

We consider networks with dynamic routing of messages. In such networks the server to which a message is directed depends on the network's state at the message arrival moment. One of the problems arising here is the analysis of probability of large fluctuations, for example probability of large delays.

There are many works where large fluctuations in networks with dynamic routing have been investigated. In [1] - [9] the networks with two servers and three independent input flows have been considered where only one flow is divided between two servers depending either on the workload of servers or on the queue lengths. In [10] a network with a group of servers and several flows has been considered where each flow is assigned to some subgroup of servers; upon its arrival a message selects a server with the shortest queue (i.e. a queue with least number of messages). In this work the large deviation principle for the flows upon the servers is proved. We would like to stress that the service time of messages in [10] is exponentially distributed. That allows to use Markov property for the flows to servers (after splitting of the input flows.)

The more full list of references can be found in the mentioned works.

Here we consider circle networks that are formed by  $k$  servers and  $k$  identical independent input Poisson flows. Messages of any flow are assigned to two nearby servers (see Figure 1) A message direction depends on the workloads on these two servers, namely upon its arrival a message is directed to the server with the smallest workload. The constant speed of work of each server is equal to 1, the discipline is "first in - first out" (FIFO). If a message finds the server busy it is put into a infinite buffer to wait for service. We consider the networks that work stationary. That means that with probability 1 the queues do not increase infinitely. But there may appear large fluctuations, for example during a short period one of flows may bring very large amount of work. We say that during this period the flow is overheated. Suppose for example that the flow  $f_1$  is overheated. Then the buffers of servers  $s_1$  and  $s_k$  that are assigned to this flow will contain a large amount of work. What is the behavior of other flows ?

In this paper we show that in case where  $f_1$  is overheated there exist at least two scenarios of network performance. What scenario is more probable and is realized depends on the rate of input flows. Namely, we show that there exists such value of input flow rate that in case when the arrival rates are above it the overheating of flow  $f_1$  coincides with overheating of all flows. Such behavior may be regarded as the spontaneous appearance of collective behaviour. On the other hand in case of low rate the overheating of  $f_1$  does not bring the overheating of other flows. There may exist also intermediate stages where the overheating of one flow coincides with the overheating of a number of neighboring flows. The existence of such intermediate stages depends on message length distribution of input flows. For some length distributions, for example for exponential one, the intermediate stages do not exist.

The proof of this result is based on large deviation principle of [11] (see also [12]). We do not present the details of this reduction. The details of application of large deviation principle for several other problems one can find in [11, 6,

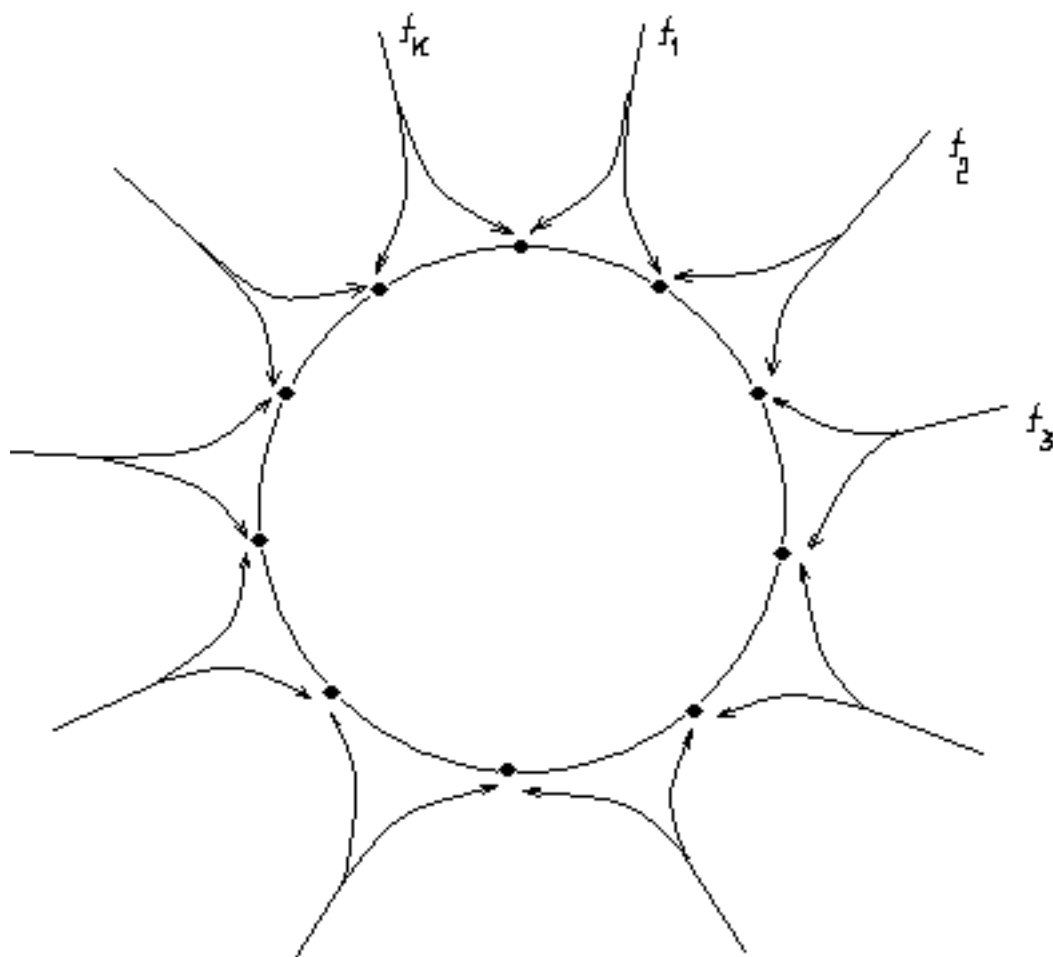


Figure 1:

7]. Instead of the proof we present some ideas needed for the proof and some fragments of rigorous arguments. Therefore the following text is not a proof in the conventional sense.

In the next section we introduce several concepts and definitions and describe the main ideas of our approach. In section 3 several examples are presented.

## 2 Cyclic networks. Main result

The cyclic network of size  $k$ ,  $k \geq 3$ , is formed by  $k$  servers and  $k$  input flows. Let  $F = \{f_1, \dots, f_k\}$  be the set of flows and  $S = \{s_1, \dots, s_k\}$  the set of servers. All servers work with speed 1, the discipline is FIFO. Each server has an infinite buffer. The messages of flow  $f_i \in F$  are served by two servers  $s_i \in S$  and  $s_{i-1} \in S$  (here always  $i \mp 1 \pmod k$ ). Each message of  $f_i$  is directed to that of servers  $s_i, s_{i-1}$  which becomes idle first.

Let the random sequence  $(\xi_n^i, \tau_n^i)$ ,  $n = \dots, -1, 0, 1, \dots$ , describes the input flow  $f_i \in F$ . The random variables  $\tau_n^i$  are the intervals between the arrivals of messages of flow  $f_i$ . The random variables  $\xi_n^i$  are the lengths of messages. We consider homogeneous Poisson flows therefore all  $(\xi_n^i, \tau_n^i)$  are independent and equally distributed,  $\tau_n^i$  are exponentially distributed with rate  $\lambda$ , i.e.  $\Pr(\tau_n^i > t) = e^{-\lambda t}$ . The distributions of input flow do not depend on  $i$ .

Introduce a random pair  $(\xi, \tau)$  that is distributed as any pair  $(\xi_n^i, \tau_n^i)$ . Each sequence  $(\xi_n^i, \tau_n^i)$  is numbered so that  $\tau_0^i < 0$  and  $\tau_1^i \geq 0$ .

We propose that  $\xi$  has exponential moments, i.e. there exist such  $\theta_+$ ,  $0 < \theta_+ \leq \infty$  that

$$\begin{aligned} \varphi(\theta) = \mathbf{E}e^{\theta\xi} &< \infty \quad \text{for } \theta < \theta_+, \\ \lim_{\theta \uparrow \theta_+} \varphi(\theta) &= \infty. \end{aligned} \tag{1}$$

The condition that guarantees the stationarity and even ergodicity is

$$\lambda\varphi'(0) < 1. \tag{2}$$

This condition is intuitively obvious but its sufficiency for the existence of stationary stage needs a proof. In particular one may construct a Liapunov function that shows that with large probability the workloads of all stations are located in a compact region. We omit this construction.

To formulate the problem that is investigated we introduce a notion of *virtual message*. That is a message that arrives with flow  $f_i$  at a nonrandom time moment, for example at 0 and has zero length. It joins a queue at  $s_i$  or  $s_{i-1}$ , depending on

where the workload is less, and waits there its service. Let  $\omega_i$  be the  $i$ th virtual message's waiting time. We are interested in the probability of large waiting for virtual message that arrived with  $f_1$ , i.e.  $\Pr(\omega_1 \geq d)$ , where  $d$  is large. In some cases, for special choice of  $\xi$  distribution the needed probability can be calculated. But in general case it is impossible to present the explicit expression. Therefore one looks for the asymptotics of probability

$$\Pr(\omega_1 > nd), \quad (3)$$

where  $n \rightarrow \infty$ . More detailed, one looks for

$$J = \lim_{n \rightarrow \infty} \frac{-1}{n} \ln \Pr(\omega_1 \geq nd). \quad (4)$$

This problem belongs to the theory of large deviation. We use this theory even where it is possible to get an explicit solution.

## 2.1 Collective fluctuations

Next theorem states the existence of input flow values  $\lambda_k \leq \lambda^k$  such that separate different types of system performance. Let

$$\hat{\lambda} = \frac{1}{\varphi'(0)}.$$

As we mentioned before the system performance is stationary if  $\lambda < \hat{\lambda}$ .

We say that the flow  $f_i$  is overheated at  $t = 0$  if  $\Gamma_i(n) = (\omega_i \geq nd)$ . It means that a virtual message that arrived with  $f_i$  is waiting for being served at least  $nd$ . Let  $\bar{\Gamma}_i(n)$  be the complement to event  $\Gamma_i(n)$ .

**Theorem 1** . *For any system size  $k$ ,  $k \geq 3$*

*and for any  $d$  there exist  $\lambda_k, \lambda^k$  (not depending on  $d$ ), such that  $0 < \lambda_k \leq \lambda^k < \hat{\lambda}$ , and*

► *if  $\lambda < \lambda_k$  then*

$$1) \Pr \left( \Gamma_1(n) \cap_{i=2}^k \bar{\Gamma}_i(n) \mid \Gamma_1(n) \right) \rightarrow 1, \text{ as } n \rightarrow \infty.$$

$$2) J = \lim_{n \rightarrow \infty} \frac{-1}{n} \ln \Pr \left( \Gamma_1(n) \cap_{i=2}^k \bar{\Gamma}_i(n) \right) = 2\theta(\lambda, 1)d$$

*where  $\theta(\lambda, 1)$  is a positive root to equation*

$$2\theta = \lambda[\varphi(\theta) - 1], \quad (5)$$

► if  $\lambda > \lambda^k$  then

$$3) \Pr \left( \bigcap_{i=1}^k \Gamma_i(n) \mid \Gamma_1(n) \right) \rightarrow 1 \text{ as } n \rightarrow \infty,$$

$$4) J = \lim_{n \rightarrow \infty} \frac{-1}{n} \ln \Pr \left( \bigcap_{i=1}^k \Gamma_i(n) \right) = k\theta^*d,$$

where  $\theta^*$  is a positive root to equation

$$\theta = \lambda[\varphi(\theta) - 1]. \quad (6)$$

It follows from this statement that in case  $\lambda < \lambda_k$  the main contribution to the probability of event  $\Gamma_1(n) = (\omega_1 > nd)$  brings the flow  $f_1$  which the virtual message arrives with. Only this flow is overheated, the others stay not overheated. But in case  $\lambda > \lambda^k$  all input flows are overheated. Though the virtual message is combined with flow  $f_1$  only its large delay is connected with an effect of collective behavior that is similar to the effect of spontaneous magnetization in statistical physics systems.

It is possible to present a more accurate statement. In the next section we introduce several definitions and present a precise theorem.

## 2.2 Random processes connected with the system and large deviations

Below we introduce several random processes connected with the cyclic network.

### 1. Process

$$\zeta^i(t) = \sum_{j: \sum_{r=1}^j \tau_r^i < t} \xi_j^i,$$

describes the amount of work brought by  $f_i$  during time interval  $[0, t)$ .

2.  $k$  dimensional Markov process  $w^1, \dots, w^k$  defined by generator

$$\begin{aligned}
Lg(w^{(1)}, \dots, w^{(k)}) = & \\
= & - \sum_{i=1}^k \frac{\partial g}{\partial w^{(i)}} \mathbb{I}_{(w^{(i)} > 0)}(w^{(i)}) + \\
& + \lambda \sum_{i=1}^k \left[ \mathbf{E}g(w^1, \dots, w^i + \xi, \dots, w^k) - \right. \\
& \quad \left. - g(w^1, \dots, w^i, \dots, w^k) \right] \mathbb{I}_{(w^i \leq w^{i+1})}(w^i, w^{i+1}) + \\
& + \lambda \sum_{i=1}^k \left[ \mathbf{E}g(w^{(1)}, \dots, w^{i+1} + \xi, \dots, w^k) - \right. \\
& \quad \left. - g(w^1, \dots, w^{i+1}, \dots, w^k) \right] \mathbb{I}_{(w^i > w^{i+1})}(w^i, w^{i+1}),
\end{aligned}$$

describes the loading of all servers. We propose that  $w^i(0) = 0$ ,  $i = 1, \dots, k$ . The symbol  $\mathbb{I}_A$  is the indicator of a set  $A$ ,  $g$  is arbitrary differentiable function in  $L_2(\mathbb{R}^k)$ . The second and third sums of the right hand side indicate that a message of length  $\xi$  arriving with  $f_{i+1}$  is directed to one of servers  $s_i$ ,  $s_{i+1}$  depending on where the loads  $w^i$ ,  $w^{i+1}$  is less.

A typical trajectory of this process is a  $k$  dimensional function with non negative components. Each component is formed by jumps and piecewise linear functions between the jumps. On the intervals where the trajectory is differentiable its derivative is either  $-1$  (if it is positive) or  $0$  (if it equals  $0$ ).

3. Process  $(\hat{w}^1(t), \dots, \hat{w}^k(t))$  describes the amount of work that was brought to each of  $k$  servers presuming that  $w^i(0) = 0$ . Let  $t_r^i > 0$ ,  $r = 1, \dots$  be the set of jumps of process  $w^i$  and  $\sigma_r^i$ ,  $r = 1, \dots$  the sizes of these jumps. Then

$$\hat{w}^i(t) = \sum_{r: t_r^i < t} \sigma_r^i. \quad (7)$$

The trajectories of this process are step functions, where jumps coincide in size and time with those of process  $w^1, \dots, w^k$ .

We define these three  $k$ -dimensional processes  $\zeta = (\zeta^1, \dots, \zeta^k)$ ,  $w = (w^1, \dots, w^k)$  and  $\hat{w} = (\hat{w}^1, \dots, \hat{w}^k)$  onto the same probability space. Here the jumps of  $w$  and  $\hat{w}$  coincide in size and time. Let  $\zeta^i$  have a jump at time moment  $t$ . Then at the same moment either  $w^{i-1}$  or  $w^i$  has a jump of the same size, depending on whether

an inequality  $w^{i-1}(t-0) < w^i(t-0)$  or the opposite one takes place. In case of equality we propose that process  $w^{i-1}$  has a jump.

Defining the processes on one probability space one can define a mapping

$$G : \mathcal{X}^o \rightarrow \mathcal{X}^o \quad (8)$$

of the set realization of processes  $\zeta$  onto the set realization of processes  $\hat{w}$ . Here  $\mathcal{X}^o$  is the set of nondecreasing stepwise functions that are equal to 0 at 0.

All event that are considered here are connected with asymptotic characteristics of network as  $n \rightarrow \infty$ . Therefore we introduce the scaled versions of processes and events.

We start with introduction of scaled versions of the processes 1 and 3:

1.

$$\zeta_n^i(t) = \frac{1}{n} \zeta^i(nt).$$

2.

$$\hat{w}_n^i(t) = \frac{1}{n} \hat{w}^i(nt).$$

Being interested in asymptotic behavior of probabilities of different events we introduce also several notions defined by the sequences of events.

Let  $\{x_{T,a}(\cdot)\}$  be a piecewise linear trajectories

$$\dot{x}_{T,a}(t) = \begin{cases} a & \text{if } t \in [0, T] \\ \lambda\varphi'(0) & \text{if } t > T \end{cases} \quad (9)$$

We say that a flow  $f_i$  is overheated on an interval  $[0, T]$  if following event takes place

$$\mathcal{A}_{\varepsilon,n}^i(x) = \left( \sup_{t \in [0, T]} |x(t) - \zeta_n^i(t)| < \varepsilon \right), \quad x = x_{T,a}, \quad a > 1, \varepsilon > 0.$$

Remark that the notion of overheated flow is introduced only for piecewise linear trajectories. We do not need more general definition.

We say that a flow  $f_i$  is not overheated on an interval  $(0, T)$  in case of event  $\mathcal{A}_{\varepsilon,n}^i(x) = \left( \sup_{t \in [0, T]} |x(t) - \zeta_n^i(t)| < \varepsilon \right)$ ,  $x = x_{T,a}$ ,  $a \leq 1$ ,  $\varepsilon > 0$ . If  $a = a^{(0)} = \lambda\varphi'(0)$  then not overheated input flow is close to its mean value.

Remark that the distribution of  $\omega_1$  coincides with the distribution of  $\min \left\{ \sup_{t \geq 0} \{\hat{w}^1(t) - t\}, \sup_{t \geq 0} \{\hat{w}^k(t) - t\} \right\}$ . That can be shown easily, we omit the needed construction. (For the problems considered in [11, 6, 7] similar facts



are explained in these works, the needed construction for a one-channel systems one can also find in monograph [15] ).

The probability (3) can be expressed in terms of scaled processes  $\widehat{w}_n^i(t)$ . It is equal to

$$\Pr(\omega_1 > nd) = \Pr\left(\widehat{\mathcal{M}}_n\right), \quad (10)$$

where event

$$\widehat{\mathcal{M}}_n = \left( \min \left\{ \sup_{t \geq 0} \left\{ \widehat{w}_n^1(t) - t \right\}, \sup_{t \geq 0} \left\{ \widehat{w}_n^k(t) - t \right\} \right\} > d \right)$$

means that both processes  $\widehat{w}_n^1$   $\widehat{w}_n^k$  intersect the line  $u(t) = d + t$ .

Using the mapping  $G$  of process  $\zeta_n$  realizations onto process  $\widehat{w}_n$  realizations (see (8)) we can introduce an event  $\mathcal{M}_n = G^{-1}\widehat{\mathcal{M}}_n$ .

**Theorem 2 .** *For any size  $k$ ,  $k \geq 3$  of the network and for any  $d > 0$  there exist not depending on  $d$   $\lambda_k$  and  $\lambda^k$ ,  $0 < \lambda_k \leq \lambda^k < \widehat{\lambda}$ , such that*

► *if  $\lambda < \lambda_k$  then there exist  $T_1^*$  and  $a_1^*$  such that*

$$\begin{aligned} 1) \quad & \lim_{\varepsilon \rightarrow 0} \lim_{n \rightarrow 0} \Pr \left( \mathcal{A}_{\varepsilon, n}^1(x_{T_1^*, a_1^*}) \cap_{i=2}^k \mathcal{A}_{\varepsilon, n}^i(x_{T_1^*, a^{(0)}}) \middle| \mathcal{M}_n \right) \rightarrow 1 \\ 2) \quad & J = \lim_{\varepsilon \rightarrow 0} \lim_{n \rightarrow 0} \frac{-1}{n} \ln \Pr \left( \mathcal{A}_{\varepsilon, n}^1(x_{T_1^*, a_1^*}) \cap_{i=2}^k \mathcal{A}_{\varepsilon, n}^i(x_{T_1^*, a^{(0)}}) \right) = \\ & 2\theta(\lambda, 1)d, \\ & \text{where } \theta(\lambda, 1) \text{ is the positive root to equation} \end{aligned}$$

$$2\theta = \lambda[\varphi(\theta) - 1], \quad (11)$$

3) *besides*

$$a_1^* = \lambda \varphi'(\theta(\lambda, 1)), \quad T_1^* = \frac{d}{\lambda \varphi'(\theta(\lambda, 1)) - 1}. \quad (12)$$

► *if  $\lambda > \lambda^k$  then there exist  $T_k^*$  and  $a_k^*$  such that*

$$\begin{aligned} 4) \quad & \lim_{\varepsilon \rightarrow 0} \lim_{n \rightarrow 0} \Pr \left( \bigcap_{i=1}^k \mathcal{A}_{\varepsilon, n}^i(x_{T_k^*, a_k^*}) \middle| \mathcal{M}_n \right) \rightarrow 1 \\ 5) \quad & J = \lim_{\varepsilon \rightarrow 0} \lim_{n \rightarrow 0} \frac{-1}{n} \ln \Pr \left( \bigcap_{i=1}^k \mathcal{A}_{\varepsilon, n}^i(x_{T_k^*, a_k^*}) \right) = k\theta^*d, \\ & \text{where } \theta^* \text{ is the positive root to equation} \end{aligned}$$

$$\theta = \lambda[\varphi(\theta) - 1], \quad (13)$$

6) besides

$$a_k^* = \lambda \varphi'(\theta^*), \quad T_k^* = \frac{d}{\lambda \varphi'(\theta^*) - 1}. \quad (14)$$

This more precise form of the theorem presents the mean dynamic of conditional process under the condition of event  $\mathcal{M}_n$ . Namely, the trajectory  $(x_{T_1^*, a_1^*}, a^{(0)}t, \dots, a^{(0)}t)$  is the conditional mean dynamic of processes  $(\zeta_n^1, \dots, \zeta_n^k)$  on the interval  $[0, T_1^*]$  in case  $\lambda < \lambda_k$ , and the trajectory  $(x_{T_k^*, a_k^*}, \dots, x_{T_k^*, a_k^*})$  is the conditional mean dynamic on the interval  $[0, T_k^*]$  in case  $\lambda > \lambda^k$ .

## 2.3 Ideas of proof

The value of  $J$  could be found if the large deviation principle for processes  $\hat{w}^1(t)$  and  $\hat{w}^k(t)$  would be known. These processes are not Poisson. But they are the functionals of  $\zeta^1(t), \dots, \zeta^k(t)$ . We remind that they are defined on the same probability space. The processes  $\zeta^i, i = 1, \dots, k$  are Poisson. Therefore one can use some known results on large deviations of Poisson processes. We use the results on large deviation principle from [11] (see also [12]).

For the problems that are considered here the application of large deviation principle consists of two parts. First one has to check the validity of the large deviation principle and describe the corresponding to the problem event. After that one has to find minimum of rate function on the event. As usual, application of large deviation principle reduces the problem to the search of the point (sometimes several points) of the event where the rate function take minimal values. A small neighborhood of the point of the minimum brings the main contribution to the asymptotic of logarithm of event probability. In our case the event is some set of trajectories and the rate function is an integral functional on the trajectories. Therefore a search of trajectory that minimizes the rate function is reduced to a variational problem.

To check the validity of large deviation principle one has to present the topological space, the sequence of measures on this space that tend to a  $\delta$ -measure and the rate function.

The topological space is a set of non decreasing  $k$  dimensional functions on  $[0, \infty)$

$$\hat{\mathcal{X}} = \mathfrak{X}^k = \{(x^1(\cdot), \dots, x^k(\cdot))\},$$

equipped with *uniformly weak* topology.

The detailed description of these functions one can find in [11], where the uniformly weak topology is introduced on this space. We use this topology. It is

explained in the same paper why the weak topology is too weak and one has to use the uniformly weak topology for the problems of the considered kind.

The sequence of measures  $P_n$  is defined by processes  $\zeta_n$ .

Finally, the rate function  $\hat{I}$  is defined on  $\langle \hat{\mathcal{X}}, P_n \rangle$  by equalities

$$\hat{I}(x^1, \dots, x^k) = \sum_{i=1}^k I(x^i), \quad (15)$$

where

$$I(x^i) = \int_0^\infty \sup_{\theta < \theta_+} \{ \theta \dot{x}^i(t) - \lambda[\varphi(\theta) - 1] \} dt, \quad (16)$$

if  $x^i$  are the absolutely continuous functions. We do not give a definition of functional on not absolutely continuous functions because condition (1) permits to avoid them.

The event  $\mathcal{F}$  where minimum of rate function we have to find is defined by the variational problem presented below.

First let us extend the mapping (8) onto the whole set  $\hat{\mathcal{X}}$  of not decreasing functions

$$G : \hat{\mathcal{X}} \rightarrow \hat{\mathcal{X}}. \quad (17)$$

The extension is denoted by the same symbol.

Consider  $X = (x^1, \dots, x^k) \in \hat{\mathcal{X}}$  and the functions  $\alpha^i(t) \in [0, 1]$ ,  $i = 1, \dots, k$ . Let

$$y^i(t) = \alpha^i(t)x^i(t) - (1 - \alpha^{i+1}(t))x^{i+1}(t), \quad i = 1, \dots, k.$$

Here  $Y = GX = (y^1, \dots, y^k)$  is a solution to the following optimization problem

$$E = \min_{\bar{\alpha}} \sum_i \int |y^i(t) - y^{i+1}(t)| dt, \quad (18)$$

$$\bar{\alpha} = (\alpha^1(t), \dots, \alpha^k(t)).$$

The event is

$$\mathcal{F} = \left\{ X = (x^1, \dots, x^k) : \min \left\{ \sup_{t \geq 0} \{y^1(t) - t\}, \sup_{t \geq 0} \{y^k(t) - t\} \right\} > d, Y = GX \right\}. \quad (19)$$

It follows from the routing rules that the flows  $(y^i)$  upon the servers given the input flows  $(x^i)$  are defined by solution to (18). That is the consequence of the system routing rules. We omit the proof of this fact.

To prove the theorem one needs to find the minimum

$$\hat{I}(\mathcal{F}) = \min \left\{ \hat{I}(X) : X \in \mathcal{F} \right\}, \quad (20)$$

As the rate function is an integral functional on  $\widehat{\mathcal{X}}$ , the solution to (20) is reduced to the solution of variational problem on the set of non-decreasing functions from  $\mathcal{F}$ .

In fact it is sufficient to find (20) on the piecewise linear functions of form (9). That is because: 1) condition (1) permits to restrict oneself by absolutely continuous functions; and 2) for the homogeneous processes with independent increments the rate function on the trajectories that connect two fixed points during a given period takes its minimum on the line that connects these points. We omit the detailed explanation of these facts (see [14, 13]).

Below we use the following not formal terminology. The trajectory  $X = (x^1, \dots, x^k)$  is called *an input flow* and the image  $Y = GX = (y^1, \dots, y^k)$  a *load flow*. A small neighborhood of trajectory  $X$  contents "the real" (jump-wise) trajectories of input processes  $\zeta_n^i$ , and a small neighborhood of trajectory  $Y$  contents "the real" trajectories of load processes  $\hat{w}_n^i$ .

A trajectory  $X = (x^1, \dots, x^k) \in \mathcal{X}$  that represents  $k$  input flows will be called a *input configuration* or simpler a *configuration*. We consider only trajectories of form (9), each is characterized by a pair  $(a^i, T^i)$  where  $a^i = \dot{x}^i(t)$  on the interval  $[0, T^i]$ . We denote by  $X = \{\bar{a}, T\}$  a configuration where all  $T^i$  are equal, here  $\bar{a} = (a^1, \dots, a^k)$ .

For functions of form (9) a solution to (18) is equivalent to solution to following optimization problem

$$\text{O1} \quad D = \min_{\bar{\alpha}} \sum_{i=1}^k |b^i - b^{i+1}|,$$

where

$$b^i = \alpha^i a^i + (1 - \alpha^{i+1}) a^{i+1}, \\ \alpha^i \in [0, 1].$$

$$\bar{\alpha} = (\alpha^1, \dots, \alpha^k) \ (i+1 \bmod k).$$

Obviously the solution exists.

In our problem for given input flow configuration  $X = \{\bar{a}, T\}$  vector  $\bar{b}$  represents the load configuration  $Y = \{\bar{b}, T\}$ .

In addition to the above correspondence of all  $X$  with  $Y$  we consider also the correspondence of subset of input flows with a subset of load flows to the assigned servers.

We are interested only in *connected* subsets  $F' = \{f_{r+1}, \dots, f_{r+l}\} \subseteq F$  of input flows. Let  $X' = \{x^{r+1}, \dots, x^{r+l}\}$  be a configuration of such flows. It is supposed that  $x^i \in X'$  are of form (9), all  $T$  are equal, i.e.  $X' = \{\bar{a}_{r+1, r+l}, T\}$  where  $\bar{a}_{r+1, r+l} = (a^{r+1}, \dots, a^{r+l}) \in \mathbb{R}_+^l$  and  $a^i = \dot{x}^i(t), t \leq T$ . The corresponding configuration of load flows to the assigned servers  $S(X') = \{s_r, s_{r+1}, \dots, s_{r+l}\}$  is

$Y' = \{\bar{b}_{r,r+l}, T\} = \{(b^r, b^{r+1}, \dots, b^{r+l}), T\}$ . Here vector  $\bar{b}_{r,r+l} \in \mathbb{R}^{l+1}$  is defined by a solution to optimization problem

**O2**

$$D_{r,r+l} = \min_{\bar{\alpha}_{r,r+l}} \sum_{i=r+1}^{r+l} |b^{i-1} - b^i|,$$

where

$$b^i = \begin{cases} (1 - \alpha^{i+1})a^{i+1} & \text{if } i = r, \\ (1 - \alpha^{i+1})a^{i+1} + \alpha^i a^i & \text{if } r < i \leq r + l - 1, \\ \alpha^i a^i & \text{if } i = r + l, \end{cases}$$

$$\alpha^i \in [0, 1],$$

$$\bar{\alpha}_{r,r+l} = (\alpha^r, \dots, \alpha^{r+l}).$$

We look now for solutions to **O2** in case  $D_{r,r+l} = 0$ . Below the notation  $D_{X,F'}$  is used to indicate that  $D$  is calculated with respect to a set of input flows  $F'$  for a given configuration of all flows  $X$ . A connected set  $F'$  is called *balanced* if  $D_{X,F'} = 0$ .

As we are interested in large deviations keeping in mind  $f_1$  only subsets  $F' \subseteq F$  with  $f_1 \in F'$  are considered.

A balanced  $F' \subseteq F$  is said to be *maximal* if  $D_{X,F''} > 0$  for any  $F'' \supset F'$ . A configuration  $X = \{\bar{a}, T\}$  may posses several maximal balanced subsets.

Our aim is to find for a configuration  $X = \{\bar{a}, T\} \in \mathcal{F}$  for which the rate function is minimal.

The rate function for  $X = \{\bar{a}, T\}$  is equal to a sum

$$I(X) = T \sum_{i=1}^k (\theta_i a^i - \lambda[\varphi(\theta_i) - 1]), \quad (21)$$

where  $\theta_i$  are defined by  $a^i = \lambda\varphi'(\theta_i)$ .

Let  $X = \{\bar{a}, T\}$  be a configuration,  $F'$  a subset of input flows and  $\mathcal{Z}$  a set of configurations such that

$$\mathcal{Z}(F') = \{Z = \{\bar{c}, T\} : c^i = a^i, \text{ if } f_i \in F'\}.$$

Rate function for this set is

$$I(\mathcal{Z}(F')) = I(X_0) = T \sum_{i: f_i \in F'} (\theta_i a^i - \lambda[\varphi(\theta_i) - 1]). \quad (22)$$

where  $X_0 = \{\bar{a}, T\}$  is a configuration with  $a_0^i = a_i$  if  $f_i \in F'$  and  $a_0^j = a^{(0)}$  if  $f_j \notin F'$

For balanced with respect to  $X = \{\bar{a}, T\}$  subset  $F'$  we consider

$$h_{X,F'} = \frac{\sum_{i: f_i \in F'} a^i}{|F'|}.$$

Here  $|F'|$  is the number of flows in  $F'$ .

Let  $h > 1$  and let  $\mathcal{Z}(h, F') = \{Z = (\bar{a}, T), h_{Z,F'} = h\}$  be a set of trajectories with fixed mean value  $h$  on  $F'$ .

**Lemma 1** *For a set of configurations  $\mathcal{Z}(h, F')$  a rate function  $I(\mathcal{Z}(h, F'))$  on interval  $[0, T]$  is equal to*

$$I(\mathcal{Z}(h, F')) = \left( h\tilde{\theta} - \lambda[\varphi(\tilde{\theta}) - 1] \right) T|F'|, \quad (23)$$

where  $\tilde{\theta}$  is determined by equality

$$h = \lambda\varphi'(\tilde{\theta}). \quad (24)$$

If  $F'$  is a maximal balanced set then

$$\lim_{\varepsilon \rightarrow 0} \lim_{n \rightarrow \infty} \frac{-1}{n} \ln \Pr \left( \mathcal{A}_{\varepsilon,n}(h, F') \right) = I(\mathcal{Z}(h, F')). \quad (25)$$

Here we denote

$$\mathcal{A}_{\varepsilon,n}(h, F') = \bigcup_{Z \in \mathcal{Z}(h, F')} \left[ \bigcap_{i: f_i \in F'} \mathcal{A}_{\varepsilon,n}^i(z_{T,c^i}^i) \cap \bigcap_{j: f_j \in F \setminus F'} \mathcal{A}_{\varepsilon,n}^j(z_{T,a^{(0)}}^j) \right],$$

$$Z = \{\bar{c}, T\} = \{z^1, \dots, z^k\}.$$

This Lemma shows that equal overheating in all flows from  $F'$  is "more" probable than the not equal one. We write "more" because that is an asymptotic result. In fact the considered probabilities decay exponentially and the exponent determined by rate function  $I$  is minimal in case of equal flows.

**Proof.** Rewrite the expression for (21)

$$I(\mathcal{Z}(h, F')) = \sum_{i: f_i \in F'} ((h + g_i)\theta_i - \lambda[\varphi(\theta_i) - 1])T,$$

where  $g_i = c^i - h$ . The solution to system  $\frac{\partial I(\mathcal{Z})}{\partial g_i} = 0$  is:  $\theta_i = \theta_j$  for all  $i, j : f_i, f_j \in F'$ . The solution is unique thanks to monotonicity in  $\theta$  of  $\frac{\partial(c\theta - \lambda(\varphi(\theta) - 1))}{\partial \theta}$ ,  $c = \lambda\varphi'(\theta)$ . Equality (25) follows from (22).

▲

It follows from  $X \in \mathcal{F}$  that

$$bT - T \geq d,$$

where  $b = \min\{b^k, b^1\}$  and  $\bar{b}$  is a solution to **O1**.

Introduce now a set of configurations  $\mathcal{C}(h, F')$  where  $F'$  is a single maximal balanced set for  $Z = \{\bar{c}, T\} \in \mathcal{C}(h, F')$ ,  $c^i = h$ ,  $f_i \in F'$ ,  $h > 1$ . Let

$$\mathcal{C} = \bigcup_{F' \subseteq F} \bigcup_{h > 1} \bigcup_{T > 0} \mathcal{C}(h, F'). \quad (26)$$

We want to find the rate function for  $\mathcal{C}$ . To find  $I(\mathcal{C})$  as  $\lambda$  is fixed one has to minimize  $\mathcal{C}$  in  $T$ ,  $h$  and  $F'$  (see (23), (24)). Because of circular structure of the network instead of minimization in  $F'$  one can minimize in number  $l = |F'|$  of input flows  $F'$ . If  $S'$  is the set of servers assigned to  $F'$  then  $|S'| = l + 1$  as  $l < k$  and  $|S'| = k$  as  $l = k$ .

Below the calculations are based on the following argument: the overload of flows  $F'$  that bring the overload of servers  $S'$  can be considered in case  $l < k$  as a overload of a one-channel system with a server of speed  $l + 1$  and input flow of rate  $\lambda l$ . In case  $l = k$  a one-channel system has a server speed  $k$  and the input flow rate  $\lambda k$ . We call such a one-channel system an *auxiliary* system.

Consider  $\mathcal{C}(F') = \bigcup_{h > 1} \bigcup_{T > 0} \mathcal{C}(h, F')$  and look for  $I(\mathcal{C}(F'))$  as  $|F'|$  is fixed.

Suppose first that  $l = |F'| < k - 1$ . Obviously  $F'$  is assigned to  $|S'| = l + 1$  servers. We can consider only such load flows that form configuration  $Y = \{\bar{b}, T\}$ , with  $b^i = b = \frac{l}{l+1}h$ ,  $s_i \in S'$ . This configuration and its small neighborhood belongs to the manifold  $\mathcal{F}$  if

$$\frac{l}{l+1}hT - T \geq d$$

Now by (23) and (24) we get

$$I(\mathcal{C}(h, F')) = \inf_T \left( \tilde{\theta} \frac{(d+T)(l+1)}{Tl} - \lambda[\varphi(\tilde{\theta}) - 1] \right) lT \quad (27)$$

where  $\tilde{\theta}$  is defined by  $\frac{l(d+T)}{T(l+1)} = \lambda\varphi'(\tilde{\theta})$ . It is easy to see that (27) has its infimum as

$$(l+1)\tilde{\theta} = l\lambda[\varphi(\tilde{\theta}) - 1]. \quad (28)$$

Denote by  $\theta(\lambda, l)$  a positive solution to the last equation. Then we get that

$$J(\lambda, l) = \inf_T I(\mathcal{C}(F')) = (l+1)\theta(\lambda, l)d \quad (29)$$

and the optimal  $b$  and  $T$  are

$$b = b(l) = \lambda \frac{l}{l+1} \varphi'(\theta(\lambda, l)), \quad T = T(l) = \frac{d}{\lambda \frac{l}{l+1} \varphi'(\theta(\lambda, l)) - 1}. \quad (30)$$

In case where  $l = k$  the sum of servers speed is  $k$  and the sum of input flow rates is  $\lambda k$ . The configuration  $Y = \{\bar{b}, T\}$  of flows to the servers is such that  $b^i = b = h$  for  $s_i \in S$  and we have the inequality  $hT - T \geq d$ . Therefore (27) becomes

$$I(\mathcal{C}(h, F')) = Tk \left( \tilde{\theta} \frac{T+d}{T} - \lambda[\varphi(\tilde{\theta}) - 1] \right).$$

Optimization in  $T$  and  $h$  gives

$$J(\lambda, k) = \inf_{T, h} I(\mathcal{C}(h, F')) = k\theta^*(\lambda)d, \quad (31)$$

where  $\theta^*(\lambda)$  is a positive root to

$$\tilde{\theta} = \lambda[\varphi(\tilde{\theta}) - 1]. \quad (32)$$

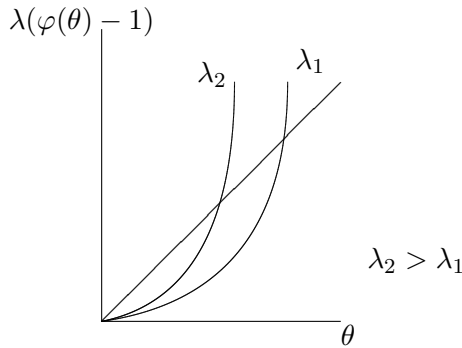
The optimal  $b$  and  $T$  are

$$b = b^* = \lambda \varphi'(\theta^*(\lambda)), \quad T = T^* = \frac{d}{\lambda \varphi'(\theta^*(\lambda)) - 1} \quad (33)$$

Remark that  $J(\lambda, k) < J(\lambda, k-1)$  for any  $\lambda$ . Really, if  $|L| = l = k-1$  then all servers are overloaded,  $J(\lambda, k-1) = k\theta(\lambda, k)d$ , where  $\theta(\lambda, k-1)$  is a solution to  $\theta = \lambda \frac{k-1}{k} [\varphi(\theta) - 1]$ . At the same time  $J(\lambda, k) = k\theta^*(\lambda)d$ , where  $\theta^*(\lambda)$  is a solution to (32). Therefore  $\theta^*(\lambda) < \theta(\lambda, k-1)$ , see Fig.2.

To get rate function  $I(\mathcal{C})$  as  $k$  and  $\lambda$  are fixed we have to find  $l$ ,  $l \leq k$ , that brings

$$\min \left[ k\theta^*(\lambda), \min_{1 \leq l < k} (l+1)\theta(\lambda, l) \right].$$



**Fig. 2.**



**Lemma 2** 1) For any  $k, l, 1 \leq l \leq k-1$  there exist such  $\lambda_{k,l}^*, 0 < \lambda_{k,l}^* < \hat{\lambda}$  that  $J(\lambda, l) < J(\lambda, k)$  as  $\lambda < \lambda_{k,l}^*$  and  $J(\lambda, l) > J(\lambda, k)$  as  $\lambda > \lambda_{k,l}^*$ .

2) For any  $l_1, l_2, 1 \leq l_1 < l_2$  there exist such  $\lambda_{l_2,l_1}, 0 < \lambda_{l_2,l_1} < \frac{l_2+1}{l_2}\hat{\lambda}$  that  $J(\lambda, l_1) < J(\lambda, l_2)$  as  $\lambda < \lambda_{l_2,l_1}$  and  $J(\lambda, l_1) > J(\lambda, l_2)$  as  $\lambda > \lambda_{l_2,l_1}$ .

**Proof.** 1). For the start we find such  $\lambda = \lambda_{k,l}^*, l \leq k-1$  that  $J(\lambda, l) = J(\lambda, k)$ . Denote  $(l+1)\theta = \vartheta_l, l < k, k\theta = \vartheta_k^*$ . It follows from (32) and (28) that

$$\vartheta_k^* = \lambda k \left[ \varphi \left( \frac{\vartheta_k^*}{k} \right) - 1 \right], \quad \vartheta_l = l \lambda \left[ \varphi \left( \frac{\vartheta_l}{l+1} \right) - 1 \right]. \quad (34)$$

By (29) and (31) the needed equality is achieved as  $\vartheta_l = \vartheta_k^*$ . Let us show that equation

$$l \left[ \varphi \left( \frac{\vartheta}{l+1} \right) - 1 \right] = k \left[ \varphi \left( \frac{\vartheta}{k} \right) - 1 \right] \quad (35)$$

has a unique solution  $\vartheta_{k,l}^*$  and

$$0 < \lambda_{k,l}^* = \frac{\vartheta_{k,l}^*}{k(\varphi(\vartheta_{k,l}^*/k) - 1)} < \hat{\lambda}.$$

Really,  $\varphi(0) = 1$ , therefore

$$\lim_{\vartheta \rightarrow 0} \frac{l \left[ \varphi \left( \frac{\vartheta}{l+1} \right) - 1 \right]}{k \left[ \varphi \left( \frac{\vartheta}{k} \right) - 1 \right]} = \frac{l}{l+1} < 1.$$

Further, if  $\theta_+ < \infty$  then

$$\lim_{\vartheta/(l+1) \rightarrow \theta_+} \varphi \left( \frac{\vartheta}{l+1} \right) = \infty, \quad \lim_{\vartheta/(l+1) \rightarrow \theta_+} \varphi \left( \frac{\vartheta}{k} \right) < \infty.$$

And if  $\theta_+ = \infty$  then

$$\lim_{\vartheta \rightarrow \infty} \frac{\varphi \left( \frac{\vartheta}{l+1} \right)}{\varphi \left( \frac{\vartheta}{k} \right)} = \infty.$$

That means that there exists  $\vartheta = \vartheta_{k,l}^*$  such that (35) takes place. The uniqueness follows from because  $\varphi$  and all its derivatives are convex..

The function  $\theta(\lambda), 0 < \theta < \theta_+$  presented by (32) is defined for  $0 < \lambda < \hat{\lambda}$  and  $\theta(\lambda)$  monotonically decreases in  $\lambda$ , (see Fig. 2). Therefore there exists such  $\lambda_{k,l}^*$ ,

$\lambda_{k,l}^* < \hat{\lambda}$ , that corresponds to  $\vartheta = \vartheta_{k,l}^*$  and for which the conditions of Lemma are fulfilled.

2). The proof of this item follows once again from existence and uniqueness of solution  $\vartheta_{l_2,l_1}$  to

$$l_1 \left[ \varphi \left( \frac{\vartheta}{l_1 + 1} \right) - 1 \right] = l_2 \left[ \varphi \left( \frac{\vartheta}{l_2 + 1} \right) - 1 \right],$$

therefore it repeats the proof of item 1). But here we have to notice that  $\theta(\lambda, l)$ ,  $0 < \theta < \theta_+$ , presented by (28) is defined for  $\lambda < \hat{\lambda} \frac{l+1}{l}$  (in (32) we had  $\lambda < \hat{\lambda}$ ). Thus  $0 < \lambda_{l_2,l_1} < \hat{\lambda} \frac{l_2+1}{l_2}$  and in general case it may be that  $\lambda_{l_2,l_1} > \hat{\lambda}$  (see Fig. 3).  $\blacktriangle$

The proof of Theorems 1 and 2 follows from the Lemmas.

Let us set

$$\lambda_k = \min_{l < k} [\lambda_{k,l}^*, \lambda_{l,1}],$$

$$\lambda^k = \max_l \lambda_{k,l}^*.$$

By Lemma 2  $J(\lambda, 1) < \min_{1 \leq l < k} J(\lambda, l)$  as  $\lambda < \min_l \lambda_{l,1}$  and therefore for fixed  $k$  we have  $J(\lambda, 1) < \min_{1 \leq l < k} \{J(\lambda, l), J(\lambda, k)\}$  as  $\lambda < \lambda_k$ . The way we get (29) indicates that as  $\lambda < \lambda_k$  the statements 1) and 2) of the Theorem take place and the values (30) correspond to the values of item 3) of the Theorem.

Further,  $J(\lambda, k) < \min_{1 \leq l < k} J(\lambda, l)$  as  $\lambda > \lambda^k$ . The way we get (31) indicates that as  $\lambda > \lambda^k$  the statements 4) and 5) of the Theorem 2 take place and the values (33) coincide with the values of item 6) of the Theorem 2.  $\blacktriangle$

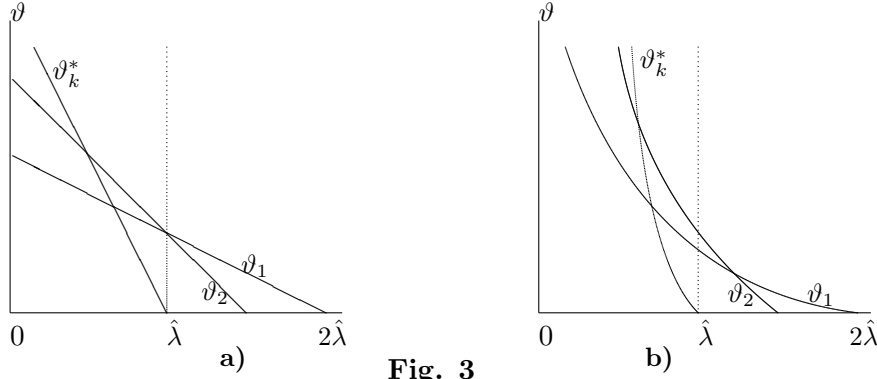
### 3 Different distributions of message length. Examples

It is clear that if  $\lambda_k = \lambda^k$  for some  $k$  then as  $f_1$  is overheated then, depending on  $\lambda$ , most probably either only  $f_1$  or all flows are overheated. The questions are: when  $\lambda_k = \lambda^k$ , what happens if  $\lambda_k < \lambda^k$ ?

To answer these questions we look at the location of curves  $\vartheta_l$  and  $\vartheta_k^*$  (see (34)) on  $(\lambda, \vartheta)$  plane.

**Proposition 1** *If  $\lambda_{l,1} > \hat{\lambda}$  for any  $1 < l < k$ , then  $\lambda_k = \lambda^k$ ,  $\lambda^k \rightarrow \hat{\lambda}$  as  $k \rightarrow \infty$ .*

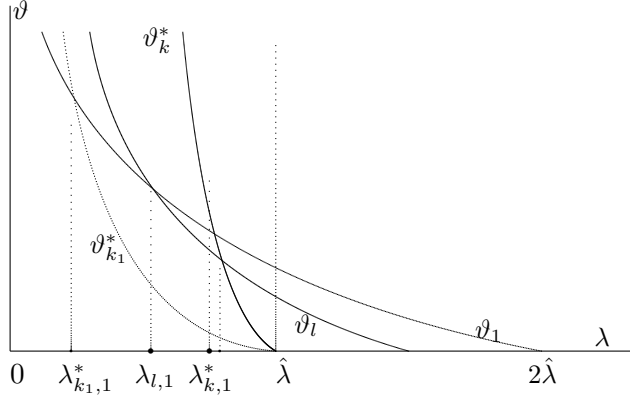
**Proof.** By Lemma 2 each pair of curves  $\vartheta_1(\lambda), \vartheta_l(\lambda)$  defined by, (34) has a unique point of intersection  $(\lambda_{l,1}, \vartheta_{l,1})$  and  $\vartheta_1 < \vartheta_l$ ,  $l > 1$ , as  $\lambda < \lambda_{l,1}$ . Therefore from  $\lambda_{l,1} > \hat{\lambda}$  follows that  $J(\lambda, 1) < J(\lambda, l)$  for all  $\lambda < \hat{\lambda}$ . Thus  $\min_{1 \leq l < k} J(\lambda, l) = \min[J(\lambda, 1), J(\lambda, k)]$ , i.e.  $\lambda_k = \lambda^k$ . The curves  $\vartheta_l(\lambda)$  do not depend on  $k$ , the curves  $\vartheta_k^*(\lambda)$  increase with  $k$ , therefore  $\lambda_{k,1}^* = \lambda^k \rightarrow \hat{\lambda}$  (see Fig.3).  $\blacktriangle$



**Fig. 3**

**Preposition 2** If  $\vartheta_1(\lambda_0) > \vartheta_l(\lambda_0)$  for some  $l$ ,  $l > 1$  and  $\lambda_0 \leq \hat{\lambda}$  then  $\lambda_k < \lambda^k$  for sufficiently large  $k$  and  $\lambda^k \rightarrow \hat{\lambda}$  as  $k \rightarrow \infty$ .

**Proof.** It follows from the condition of Lemma that  $\lambda_{l,1} \leq \hat{\lambda}$ . Remember that  $\vartheta_k^*(\lambda)$  increases in  $k$ . If  $k$  is sufficiently large then the pairs  $\vartheta_k^*, \vartheta_1$  and  $\vartheta_k^*, \vartheta_l$  intersect at points  $\lambda_{k,1}^*$  and  $\lambda_{k,l}^*$  where  $\lambda_{k,l}^* > \lambda_{l,1}$ ,  $\lambda_{k,1}^* > \lambda_{l,1}$ , and  $\lambda_{k,l}^* > \lambda_{k,1}^*$ . Therefore  $\vartheta_k^*(\lambda) > \vartheta_1(\lambda) > \vartheta_l(\lambda)$  as  $\lambda_{l,1} < \lambda < \lambda_{k,1}^*$  (see Fig. 4), i.e.  $\min_{1 < l < k} J(\lambda, l) < \min[J(\lambda, 1), J(\lambda, k)]$  for such  $\lambda$  and that means  $\lambda_k < \lambda^k$ . As in Preposition 1  $\lambda^k \rightarrow \hat{\lambda}$  as  $k \rightarrow \infty$ .  $\blacktriangle$



Here  $k > k_1$ ,  $\lambda_{k_1} = \lambda^{k_1} = \lambda_{k_1,1}^*$ ,  $\lambda_k < \lambda^k$ ,  $\vartheta_l < \vartheta_1$  as  $\lambda_{l,1} < \lambda < \lambda_{k,1}^*$

**Fig. 4**

Below we present several examples to show the realization of described scenarios.

**1. Exponential distribution of message length.**

The density of message length distribution is  $ce^{-cx}$ ,  $c > 0$  and  $\varphi(\theta) = \frac{c}{c-\theta}$ , the condition of stability is  $\lambda < \hat{\lambda} = c$ . The equations (34) are linear having the form

$$\vartheta_l = c + l(c - \lambda), \quad \vartheta_k^* = k(c - \lambda).$$

It is clear that  $\vartheta_1(\lambda) < \vartheta_l(\lambda)$  as  $\lambda < c = \hat{\lambda}$ . On  $(\lambda, \vartheta)$  plain all  $\vartheta_l(\lambda)$  intersect at a point  $(c, c)$  and  $\lambda_k = \lambda^k = c^{\frac{k-2}{k-1}}$  (See Fig 3 ).

**2. The density of message length distribution is**

$$\frac{1}{2} \left( (c+g)e^{-(c+g)x} + (c-g)e^{-(c-g)x} \right), \quad g < c.$$

Here we have  $\varphi(\theta) - 1 = \frac{\theta(c-\theta)}{(c-\theta)^2 - g^2}$ . The stability condition is  $\lambda < \hat{\lambda} = \frac{c^2 - g^2}{c}$ .

For  $\lambda \leq \hat{\lambda}$  we can find  $\vartheta_l(\lambda) = (l+1)\theta(\lambda)$  using (28)

$$(c-\theta)^2 - \frac{l}{l+1}\lambda(c-\theta) - g^2 = 0;$$

the needed expression is

$$\begin{aligned}\vartheta_l(\lambda) &= (l+1) \left[ c - \frac{\lambda l}{2(l+1)} - \sqrt{\frac{\lambda^2 l^2}{4(l+1)^2} + g^2} \right] \\ &= c + l \left[ c - \frac{\lambda}{2} - \sqrt{\frac{\lambda^2}{4} + g^2 \frac{(l+1)^2}{l^2}} \right].\end{aligned}\quad (36)$$

The value in square brackets of (36) increases in  $l$ , thus  $\vartheta_l(\lambda)$  increases in  $l$  as  $\lambda \leq \hat{\lambda}$ . That means that all  $\vartheta_l(\lambda)$  intersect as  $\lambda > \hat{\lambda}$  (see Fig 3 b). By Proposition 1 that indicates that  $\lambda_k = \lambda^k$ .

This example demonstrates the possibility of the following scenario: the overheat of  $l$  connected flows may bring not only overload of  $l+1$  assigned servers, but also overload of other servers by not overheated flows.

For example, let  $k = 3$ ,  $c = 1$ ,  $g = 0.5$ , here  $\hat{\lambda} = 0.75$ . The numerical estimation gives  $\lambda_{3,1}^* \sim 0.418$ , i.e.  $2\lambda_{3,1}^* > \hat{\lambda}$ . That means that as  $\lambda$  is inside an interval  $\left(\frac{\hat{\lambda}}{2}, \lambda_{3,1}^*\right)$  then the overheated  $f_1$  brings not only the overload of  $s_1$ ,  $s_3$  but also the overload of  $s_2$  fed by not overheated  $f_2$  and  $f_3$ . And it is easy to estimate that  $s_1$ ,  $s_3$  are overloaded with greater speed than  $s_2$ .

### 3. Constant message length.

Let  $\Pr(\xi = x) = \delta(1/c)$ ,  $c > 0$ . Here  $\varphi(\theta) = e^{\theta/c}$ . The stability condition is  $\lambda < c$ .

For  $\lambda = \hat{\lambda} = c$  we get by (34) that  $\frac{\vartheta_l(c)}{c} = l(e^{\vartheta_1(c)/c(l+1)} - 1)$ . It is easy to check that  $\frac{\vartheta_1(c)}{c} > 2$ , and  $\lim_{l \rightarrow \infty} \frac{\vartheta_l(c)}{c} = 2$ , i.e.  $\vartheta_1(\hat{\lambda}) > \vartheta_l(\hat{\lambda})$  as  $l$  is sufficiently large. By Proposition 2  $\lambda_k < \lambda^k$  as  $k$  is sufficiently large.

The numerical estimates performed for  $c = 1$  and  $k \leq 35$  indicate that the behavior of  $\lambda_k$  and  $\lambda^k$  changes as  $k$  increases (and so do also the scenario of  $\vartheta_l$  and  $\vartheta_k^*$  intersection).

a) Here  $\lambda_k = \lambda^k = \lambda_{k,1}^*$  as  $3 \leq k \leq 12$ ,  $\lambda_k$  increases in  $k$ .

For example  $\lambda_k \sim 0.311$  as  $k = 3$ ;  $\lambda_k \sim 0.667$  as  $k = 5$ ;  $\lambda_k \sim 0.857$  as  $k = 10$ ;  $\lambda_k \sim 0.883$  as  $k = 12$ .

b) As  $k > 12$  the value  $\lambda_k$  does not change and is equal to  $\lambda_{2,1} \sim 0.888$ ;  $\lambda^k$  increases in  $k$ ,  $\lambda_k < \lambda^k$ .

c) Up to  $k = 28$   $\min_l J(\lambda, l) = J(\lambda, 2)$  as  $\lambda_k < \lambda < \lambda^k$ . Therefore most probab two flows are overheated. For example  $\lambda^k \sim 0.910$  as  $k = 15$ ;  $\lambda^k \sim 0.935$  as  $k = 20$ ;  $\lambda^k \sim 0.940$  as  $k = 25$ .

d) As  $k > 29$  in addition to interval  $(\lambda_k, \lambda_{3,2})$ , where  $\min_l J(\lambda, l) = J(\lambda, 2)$ , there appears an interval  $(\lambda_{3,2}, \lambda^k)$  where  $\min_l J(\lambda, l) = J(\lambda, 3)$ , and most probable three flows are overheated. Here  $\lambda_{3,2} \sim 0.956$ ;  $\lambda^k \sim 0.959$  as  $k = 30$ ;  $\lambda^k \sim 0.965$  as  $k = 35$ .

These estimates show that nonzero interval  $(\lambda_k, \lambda^k)$  is small and  $\lambda^k$  is close to  $\hat{\lambda}$  as  $k > 12$ . Presented numerical data and some analytic investigation suggest that as  $k \rightarrow \infty$  a "jump" from  $l$  to  $k$  (where  $l = \arg \min_{1 \leq m \leq k} J(m, \lambda)$  changes its value from  $l < k$  to  $k$ ) happens at  $l = o(k)$ .

All numerical data are presented with accuracy 0.0005

We want to remark that in case of large fluctuations the collective behavior of dependent servers may take place for others, not circular networks.

## 4 Acknowledgement

N.D.V. thanks V. Blinovski, K. Duffi, S. Pirogov and Yu. Suhov for the useful discussions. The work of E.A.P. was partly supported by Grant RUM1-2693-MO-05 of CRDF.

## References

- [1] *Alanyali M., Hajek B.* On large deviations in load sharing networks // Ann. Appl. Probability. 1998, V. 8, 1, P. 67-97 .
- [2] *Turner S.R.E.* Large deviations for Join the Shortest Queue // Fields Inst. Communications. 2000, V. 28, P. 95-106.
- [3] *McDonald D.R. and Turner S.R.E.* Resource Pooling in Distributed Queueing Networks // Fields Inst. Communications. 2000, V. 28, P.107-131.
- [4] *Foley R.D., McDonald D.R.* Join the shortest queue: stability and exact asymptotics // Ann. Appl. Probab. 2001, V. 11, 3, P. 569-607.
- [5] *Pechersky E.A., Suhov Y.M., Vvedenskaya N.D.* Large deviations in a two-server system with dynamic routing // Tech. report, Isaac Newton Institute for Math. Sci.2003, preprint NI03075-IGS.

- [6] *N.D. Vvedenskaya, E.A. Pechersky, Y.M. Suhov* Large Deviations in Some Queueing Systems// Problems Inform. Transmissions, 2000, V. 36, 1, P. 42-53.
- [7] *Aspandijarov S., Pechersky E.*, One large deviations problem for compound Poisson processes in queueing theory // Markov Processes and Relat. Fields, 1997, V. 3, 3, P. 333-366.
- [8] *Duffy K., Malone D., Pechersky E., Suhov Y., Vvedenskaya N.* Large deviations provide good approximation to queueing system with dynamic routing // 2004, Tech. report, Dublin Insitute for Advanced Studies.
- [9] *Duffy K., Pechersky E.A., Suhov Y.M, Vvedenskaya N.D.* Using estimated entropy in a queueing system with dynamic routing // Markov Process and Related Fields. 2007, V.13, 1, T. 57-84.
- [10] *Puhalskii A.A., Vladimirov A.A.* A large deviation principal for join the shortest queue // Mathematics of Operation Research, V. 32, 3, P. 700-710, 2007.
- [11] *R.L.Dobrushin, E.A. Pechersky* , Large deviations for random processes with independent increments on infinite intervals, Problems Inform. Transmissions, V. 34, 4, 1998, P. 354-382.
- [12] *Li Z-H, Pechersky E.* On large deviations in queueing systems // Resenhas IME-USP 1999, V. 4, 2, P. 163-182.
- [13] *Dobrushin R.L., Pechersky E.A.* Large deviations for tandem queueing systems, Journal of Applied Mathematics and Stochastic Analysis // 7, 3, 1994, 301-330.
- [14] *Lynch J., Sethuraman J.* Lagre deviations for processes with independent increments // Ann. Prob. 1987, 15, 2, 610-627.
- [15] *Borovkov A.A.* Stachastic Processes and Queueing Theory, Springer-Verlag, 1976.